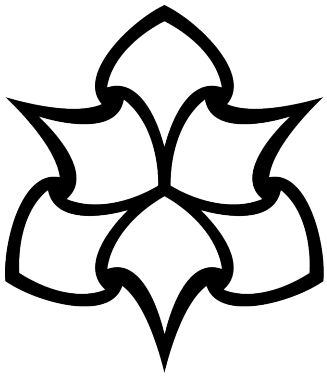

Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia.

Paul CONROY✉ *, Christina SOTIROPOULOU DROSOPOULOU*, Gina F. HUMPHREYS, Ajay D. HALAI, & Matthew A. LAMBON RALPH✉

Neuroscience & Aphasia Research Unit,
Division of Neuroscience & Experimental Psychology,
School of Biological Sciences
University of Manchester UK

✉Address correspondence to: Dr. Paul Conroy or Prof. Matt Lambon Ralph, Neuroscience and Aphasia Research Unit (NARU), School of Biological Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, UK. E-mail: paul.conroy@manchester.ac.uk or matt.lambon-ralph@manchester.ac.uk

* - Joint first authors

ORCiD:
Conroy: 0000-0003-2189-4149
Sotiropoulou Drosopoulou: 0000-0002-7237-3514
Humphreys: 0000-0002-4157-3409
Halai: 0000-0003-1725-7948
Lambon Ralph: 0000-0001-5907-2488

Running title: speed + accuracy of word retrieval in aphasia

Acknowledgements: We would like to thank the participants with aphasia who took part in this study. This research was supported by grants from the Rosetrees Foundation, the MRC (MR/J004146/1) and ERC (GAP: 670428 - BRAIN2MIND_NEUROCOMP).

Abstract

One third of stroke survivors experience deficits in word retrieval as a core characteristic of their aphasia, which is frustrating, socially-limiting and disabling for their professional and everyday lives. The, as yet undiscovered, "holy grail" of clinical practice is to establish a treatment that not only improves item naming but also generalizes to patients' connected speech. Speech production in healthy participants is a remarkable feat of cognitive processing being both rapid (at least 120 words per minute) and accurate (~one error per 1000 words). Accordingly, we tested the hypothesis that word-finding treatment will only be successful and generalize to connected speech if word retrieval is both accurate and quick.

This study compared a novel combined speed- and accuracy-focused intervention - 'repeated, increasingly-speeded production' - to standard accuracy-focused treatment. Both treatments were evaluated for naming, connected speech outcomes, and related to participants' neuropsychological and lesion profiles. Twenty participants with post-stroke chronic aphasia of varying severity and subtype took part in 12 computer-based treatment sessions over 6 weeks. Four carefully-matched word-sets were randomly allocated either to the speed- and accuracy-focused treatment, standard accuracy-only treatment, or untreated (two control sets). In the standard treatment, sound-based naming cues facilitated naming accuracy. The speed- and accuracy-focused treatment encouraged naming to become gradually quicker, aiming towards the naming time of age-matched controls.

The novel treatment was significantly more effective in improving and maintaining picture naming accuracy and speed (reduced latencies). Generalization of treated vocabulary to connected speech was significantly increased for all items relative to the baseline. The speed- and accuracy-focused treatment generated substantial and significantly greater deployment of targeted items in connected speech. These gains were maintained at one-month post intervention. There was a significant negative correlation for the speed- and accuracy-focused treatment between the patients' phonological scores and the magnitude of the therapy effect, which may have reflected the fact that the substantial, beneficial effect of the novel treatment generated a ceiling effect in the milder patients. Maintenance of the speed- and accuracy-treatment effect correlated positively with executive skills. The neural correlate analyses revealed that participants with the greatest damage to the posterior superior temporal gyrus extending into the white matter of the inferior longitudinal fasciculus, showed the greatest

speed- and accuracy treatment benefit. The novel treatment was well tolerated by participants across the range of severity and aphasia subtype, indicating that this type of intervention has considerable clinical utility and broad applicability.

Keywords: aphasia, word retrieval, speed, naming, treatment, stroke.

Introduction

Fluent speech requires rapid, errorless retrieval of vocabulary, which occurs at a rate of at least two words per second and less than one error per 1000 words (Bird et al., 2000; Levelt, 1989). Aphasia occurs in at least one third of stroke survivors (The Stroke Association (UK), 2016). Failures, errors or delays in word retrieval (anomia) are the most pervasive aphasic symptoms (Laine and Martin, 2006). Anomia treatment typically involves single-item picture naming. There is a strong clinical belief that there is a lack of generalization to connected speech for standard naming therapies (Nickels, 2002; Wisenburn and Mahoney, 2009), yet typically studies (a) have lacked a systematic method for assessing generalization and (b) have been underpowered.

Given that connected speech is highly demanding in terms of speed and accuracy, we hypothesised that retrained vocabulary will only generalise if it can be retrieved within the demanding time window required by connected speech (Conroy et al., 2009; Crerar, 2004). This hypothesis aligns with the broader observations that (a) naming speed is an important variable for both assessment and treatment tasks (McCall et al., 1997) and (b) in mild aphasia, expressive vocabulary may be largely recovered except for delayed naming latencies (Crerar, 2004).

To tackle this critical clinical need, we developed a novel treatment to reduce speed and increase picture naming accuracy, simultaneously ('repeated, increasingly-speeded production': RISP). This intervention was directly compared to (a) a standard treatment that targeted accuracy alone and (b) no treatment. We hypothesized that (i) RISP would generate greater improvements in both naming speed and accuracy, and (ii) speedier naming would increase production of treated words in connected speech (evaluated through a newly-developed, systematic method). Finally, we related the patients' variable therapy outcomes to both their background neuropsychological profiles and the distributions of the underlying lesions.

Materials and methods

Participants

The participants were recruited from a post-stroke database within the Neuroscience and Aphasia Research Unit. The database consisted of seventy participants with chronic aphasia following cerebrovascular accident (CVA). All were recruited from aphasia support groups or speech therapy services in Greater Manchester and North-West England. Participants covered the full-range of aphasia severity and multiple subtypes. All were right handed, native English speakers, who had sustained one left hemisphere stroke at least one year prior to recruitment, had normal or corrected-to-normal hearing and vision, had no co-existing neurological impairments and had no contradistinctions for MRI scanning. Nineteen participants had no contradistinctions to MRI scanning (i.e. no pacemakers, metal implants, claustrophobia, etc.), however one patient had a metal implant. This meant that neuroimaging data was collected only from nineteen patients. Neuroimaging data from a healthy age and education matched control group (8 female, 11 male) was used to determine abnormal voxels using the automated lesion identification procedure (Seghier et al., 2008). All participants gave written informed consent with approval from the local ethics committee.

From the full database, twenty participants (11 males, 9 females; mean age 65.2 years, SD = 11.7) took part in the study. Prerequisites for participating were to have minimal repetition skills (>40% on an immediate word repetition test: (Kay et al., 1996). Participants with global aphasia, severe perceptual problems or with very severe naming difficulties (below 8% or 5/60 on the Boston Naming Test: (Goodglass et al., 2000), were excluded from the study. All other levels and types were included so that the newly-developed therapy could be trialled across a full range of patients. Demographic details of the participants are given in Supplementary Table 1 together with baseline picture naming accuracy and speed, and production of the same vocabulary items in connected speech (with participants ordered according to their BNT naming accuracy).

Background assessments

Before taking part in this study, participants also completed extensive linguistic and cognitive assessment. The results are summarised in Supplementary Tables 2 and 3. The background assessment battery included the following specific tests. The Boston Naming Test (BNT)

(Goodglass et al., 2000) was used to assess word-finding difficultes. Four repetition tasks were used (from Kay et al., 1996): (a) word repetition immediate; (b) word repetition delayed; (c) non-word repetition immediate; (d) non-word repetition delayed. Two other phonological tasks included word and non-word minimal pairs (Kay et al., 1996). Participants also completed six tests of comprehension and semantic memory: (a) spoken sentence comprehension from the Comprehensive Aphasia Test (CAT: (Swinburn et al., 2004); (b) Synonym Judgement Test (Jefferies et al., 2009); and from the Cambridge Semantic Battery (Bozeat et al., 2000): (c) picture naming; (d) spoken word-to-picture matching; (d) written word-to-picture matching; and (e) the picture-version of the Camel and Cactus Test (CCT) of semantic association knowledge. To test short-term memory skills, the forward and backward memory span assessments were administered (Wechsler, 1945). Two executive tests were completed: (a) Brixton Spatial Rule Anticipation Test (Burgess and Shallice, 1997) and Raven's Coloured Progressive Matrices (Raven, 1962). Speech production deficits were assessed by coding responses to the 'Cookie theft' picture in the BDAE, which included tokens (TOK), mean length of utterance (MLU), type/token ratio (TTR) and words-per-minute (WPM). All scores were converted into percentages; if no maximum was available we used the maximum score across the participants. Following previous studies, we utilised principal component analysis (PCA; SPSS v.22) to express the underling dimensions of performance variation (Butler et al., 2014; Halai et al., 2017). A PCA with varimax rotation was calculated for these behavioural measures for our full N=70 chronic aphasia patient dataset. We performed the PCA on the full available dataset in order to: 1) maximise coverage of the multidimensional space and 2) achieve robust weighted-averages for the scores of the patients on the extracted PCA components. Four principal components with an eigenvalue>1 were extracted; these corresponded to phonological, semantic, executive and speech quanta dimensions (see Halai et al., 2017 for the details of these principal components and their lesion correlates). Patients' component scores on the four extracted components were reconstucted using regression for the entire dataset (N=70). To explore the relationship between therapy outcome and background language-cognitive skills, the component scores for the subset of 20 patients included in the therapy (one did not have a factor score as we did not have full background neuropsychological data) were correlated with their therapy outcomes (1 week vs. baseline and 1 month vs. baseline). We note here that it is preferable to compute the PCA and resultant component scores on the full patient dataset as this ensures that (a) the PCA is as robust as possible and (b) places

the scores for the therapy subgroup in relation to the full range of aphasia severity (as shown in Supplementary Figure 1).

Acquisition of Neuroimaging data

High resolution structural T1-weighted MRI scans were acquired on a 3.0 Tesla Philips Achieva scanner using an 8-element SENSE head coil. A T1-weighted inversion recovery sequence with 3D acquisition was employed, with the following parameters: TR (repetition time) = 9.0 ms, TE (echo time) = 3.93 ms, flip angle = 8°, 150 contiguous slices, slice thickness = 1 mm, acquired voxel size 1.0 x 1.0 x 1.0 mm^3, matrix size 256 x 256, FOV = 256 mm, TI (inversion time) = 1150 ms, SENSE acceleration factor 2.5, total scan acquisition time =575 s.

Analysis of Neuroimaging data

Structural MRI scans were pre-processed with Statistical Parametric Mapping software (SPM8: http://www.fil.ion.ucl.ac.uk/spm/). The images were normalised into standard Montreal Neurological Institute (MNI) space using a modified unified segmentation-normalisation procedure optimised for focal lesioned brains (Seghier et al., 2008). Data from all participants with stroke aphasia and healthy controls were entered into the segmentation-normalisation that combines segmentation, bias correction and spatial normalisation through the inversion of a single unified model (Ashburner and Friston, 2005). Each patient's lesion was identified using an outlier detection algorithm based on fuzzy clustering. The default parameters were used except we modified the U-threshold from 0.3 to 0.5 after comparing sample results to those from an expert neurologist. The images were individually, visually inspected with respect to the original scan, and were used to create the lesion overlay map in Figure 1. We note that although referred to as an automated 'lesion' segmentation method, the technique detects areas of unexpected tissue class; thus identifying missing grey and white matter but also augmented CSF space. We then smoothed the T1-weighted images (8mm full width half maximum Gaussian kernel) and created separate models where we correlated with magnitude of the RISP effect to the signal intensity for each voxel in the whole brain using a voxel-based correlational methodology (VBCM) (Tyler et al., 2005), a variant of voxel symptom lesion mapping (VSLM) (Bates et al., 2003). An additional covariate was added to each model to account for lesion volume. Overlays were thresholded at $p<0.005$ voxel height and cluster corrected at familywise error of $p<0.05$, while including additional covariates of age, years of education, months post onset and lesion volume. All anatomical labels were based on the Harvard-Oxford atlas in MNI space.

Figure 1 about here

Therapy Methods

Stimuli

One reason for the dearth of information with regard to generalization from naming therapy to connected speech, is the lack of a systematic assessment method (Maendl, 1998). In order to measure word retrieval in both picture naming and connected speech, four detailed multi-event pictures were selected (from the "Where's Wally/Waldo?" publications). These contained detailed depictions of hundreds of items and events (e.g., animals, objects and events at a busy zoo or fairground) from which a small minority of target items were selected. To assess and treat confrontational naming for these targets, we selected new pictures of the same exemplars (presenting each exemplar singly and without any background). The 120 target stimuli were all nouns, selected to meet the following criteria: (a) named spontaneously in control participants' scene descriptions by more than 3/10 participants (participants and patients were asked to describe freely and were not directed to any areas or items within the scene); (b) targets could be depicted singly in a new picture with 100% name agreement (thus pictureable nouns such as *bench* rather than *water* or actions were selected); and (c), items did not have alternative names (e.g., *dodgems* and *bumper cars*).

From these 120 nouns, four matched sets of 20 items were selected; two sets were allocated to the treatment conditions, (described below). The remaining two sets served as untreated control items (thus controlling for any non-specific effects, including the small boost in performance that can result from repeated assessment (Nickels, 2002)). One treatment set and its paired untreated set related to two of the four composite pictures. The other treatment set and its paired control related to the remaining two composite pictures. This allowed us to separate the effects of each therapy by avoiding target vocabulary for the two treatments appearing in the same composite picture. The allocation of picture sets to the two treatments was counterbalanced across participants. The word sets were matched (Van Casteren and Davis, 2007) for (a) the likelihood of retrieval in the spontaneous picture descriptions by the control participants; (b) frequency from the British National Corpus (BNC Consortium, 2007); and (c) phoneme length.

Baseline and post-therapy assessment

Baseline performance for the four composite picture descriptions and 80-item confrontational naming test were assessed twice before therapy commenced (across four separate assessment sessions with composite description assessed before the confrontational naming). There was no significant change in performance across the two baseline assessments (confirming a stable baseline) and thus we compared the post-therapy results to the first assessment. Post-therapy performance was assessed at one-week and again at one-month to establish the longer-term benefits of the therapy (no maintenance or practice regimes were used post therapy). An additional, fifth composite picture description was assessed before and after therapy. No vocabulary from this fifth picture was included in the therapy as treated or untreated items. The fifth picture acted as a control for the target composite pictures in order to control for non-specific improvements that might arise simply from repeated assessment.

For the picture naming assessment, participants were presented with all 80 items in random order. Each picture was presented simultaneously with an auditory beep and remained on the screen for a maximum of ten seconds (using E-Prime software (Schneider et al., 2002). Audacity software was used to measure naming latencies by calculating the time elapsed from the beep to the onset of the participant's correct response. When no correct name was produced, the reaction time for that trial was treated as missing data.

To elicit connected speech samples, participants were informed that they were going to see four 'busy' pictures, one at a time on a computer screen. They were asked to describe what they saw in each picture in as much detail as they could for about 5-10 minutes. Participants' responses were digitally audio-recorded. The order of presentation was randomised across participants, thus counterbalancing any effect of relative difficulty.

Treatments

Figure 2 about here

The treatments were delivered in two phases (see Figure 2), each containing two sessions per week for three weeks (six treatment sessions per phase). In the first phase, only standard therapy was administered for all items (n=40) in order to boost naming accuracy before

introducing any speed requirement. In the second phase, standard (accuracy-only) treatment was continued for one set, whilst the other was treated with RISP (see below). In both phases, stimuli in each set were randomised and the order of sets was counterbalanced across sessions. Please see Table 1 for treatment protocols for SP and RISP. Treatment sessions lasted between 30 and 50 minutes dependant on participant need for comfort breaks.

Table 1 about here

***Standard production (SP)*:** This was a standard, increasing cues, naming therapy, which aimed to improve participants' picture naming accuracy only. Participants were asked to name each picture, presented on a computer screen, in 10 seconds without support, i.e. with no cues. After each naming attempt, feedback was provided both verbally by the experimenter and presented in writing on the screen. Initially, minimal cues were provided (the initial consonant and vowel of the target word, e.g., "wi" for 'window') but the cues were increased if naming was not achieved (e.g., "wind" for 'window', and then the whole word 'window'). Participants worked through all therapy items three times per session. There were no auditory cues presented in this standard therapy to indicate any type of time pressure.

Repeated, Increasingly-Speeded Production (RISP treatment): This treatment was a hybrid intervention that combined cued naming with the deadline naming method used in experimental psycholinguistics (Hodgson and Lambon Ralph, 2008; Vitkovitch and Humphreys, 1991). Participants were instructed that the computer would present the picture for a limited time and their task was to try to name the picture before the beep at the end of the stimulus presentation. In each therapy session, the presentation duration/time-to-the-beep was reduced (see below). During each trial, the target picture was presented on the computer screen. At the end of the allotted time, the picture disappeared and a beep sound was produced by the computer. A blank screen was displayed for 1000msec. Participants were then presented with the written target word on the screen and the correct spoken name of the picture was played by the computer. Following an incorrect response, participants were asked to repeat the correct name three times. Participants cycled through all therapy items three times per session. This matched the number of item exposures between RISP and SP within each treatment session.

The naming deadline was shortened systematically across the six RISP sessions. The initial picture exposure time was set to the mean of all patients' baseline picture naming speed (3 seconds). This ensured that each participant's first 'speeded' naming attempt would feel reasonably natural. The ultimate target deadline in the 6th RISP session was 1 second, which matched the mean naming speed of elderly neurotypical participants (mean naming time: 1002 msec). The target naming speed was reduced in a systematic way: session 1 = 3 seconds, session 2 = 2.5 seconds, session 3 = 2 seconds, session 4 = 1.6 seconds, session 5 = 1.3 seconds, and session 6 = 1 second. The same target naming speed was used for the three cycles within each session and only reduced on the start of the next session. It was not necessary for participants to actually 'beat the beep'; rather the attempt to do so was expected and did reduce naming latencies over the course of the treatment.

Scoring

Participant's performance was scored based on their first response for all picture naming. Self-corrections were considered correct if the correct name was produced immediately after the first response.

Analysis of the main therapy data

For the three sets of target data (picture naming accuracy, picture naming speed, and word retrieval accuracy in the composite pictures), we carried out the same set of hierarchically-structure analyses. First, we conducted a global ANOVA with picture set (the treated and untreated items) and time (pre- vs. immediately post therapy vs. one month follow up) as main factors – which allows us to specify if there were changes in performance before and after intervention, and if these varied for treated and untreated sets. We then unpicked the nature of the significant interactions with planned ANOVA and t-tests: before and after intervention, each treated set was compared to its matched untreated set, and the two treated sets were compared to each other. Our *a priori* expectations were that performance on the therapy sets would be significantly improved after therapy and better than that observed for the untreated items. Analyses were run in SPSS v22.0.

Results

Picture naming accuracy after 1st treatment phase

In the first phase, the standard therapy (SP) was administered for all items (n=40). Naming accuracy at the end of this phase are reported in Table 2. Set A progressed to be treated with SP in the 2nd phase of treatment, and set B with RISP. The mean accuracy for set A was 78.0, and set B 81.25 (a non-significant difference: two-tailed t= -0.43, p=0.66). Thus the main study comparison ANOVAs carried out at the end of the 2nd treatment phase were not biased by the (equivalent) performance on the sets after the initial treatment phase.

Table 2 about here

Picture naming accuracy after 2nd treatment phase

A global 3×4 ANOVA was conducted with the factor of time and treatment. These analyses were concerned with the cumulative effects of SP alone (phases 1+2 – i.e., Set A) versus SP followed by RISP (phases 1+2 – i.e., Set B). The three time-points were: baseline (pre-1st phase of treatment), one week post-2nd phase of treatment, one month post-2nd phase of treatment. The four treatment conditions were: SP, RISP, untreated SP, and untreated RISP. This 3x4 ANOVA indicated that there was a main effect of time ($F(2,38) = 55.6$, $p < .0005$), a main effect of treatments ($F(3,57) = 35.7$, $p < .0005$), and a significant interaction between time and treatments ($F(6,114) = 18.0$, $p < .0005$; see Figure 3a) – indicating very different effects of therapy on the treated and untreated items.

We explored the nature of this interaction through three follow-up ANOVAs. First, we compared each treatment condition to its matched control set across the three time-points, (through two 2×3 ANOVAs where the first factor compared each treatment type to its own control: i.e., SP vs. untreated SP; RISP vs. untreated RISP). These ANOVAs showed that both therapies generated significantly improved accuracy scores relative to their control sets (significant interaction: $p < .0005$ for both therapies). For RISP, a significant interaction between 'Time Point' and 'Treatment' was found: ($F(2,38) = 34.643$, $p < .0005$, partial $\eta^2 = .65$). For SP, a similarly robust significant interaction between 'Time Point' and 'Treatment' was evident: $F(2,38) = 14.935$, $p < .0005$, partial $\eta^2 = .44$).

Direct comparison of the two treatments, through another 2×3 ANOVA (SP vs. RISP; over the three time points), indicated that there was a trend towards a borderline interaction between time and treatment: $F(2,38) = 2.3$, $p = .117$. Planned t-tests showed that both therapies significantly increased picture naming accuracy between the baseline and post-treatment assessments ($p < .0005$), and that the RISP therapy effect was significantly greater than SP not only at the 1week post-treatment assessment ($p < .0005$), but also at the follow-up (1 month) assessment ($p = .001$).

Figure 3 about here

Picture naming speed after 2nd treatment phase

Exactly the same set of planned ANOVAs and t-tests were used to examine the naming speed for correctly named items (the overall results are shown in Figure 3b). In the global 3 (time point) × 4 (picture sets) ANOVA, there was a main effect of 'Time Point': $F(2,36) = 21.1$, $p < .0005$, no main effect of 'Treatment' factor [$F(3,54) = 1.7$, $p = .174$], but a significant interaction between 'Time Point' and 'Treatment' [$F(6,108) = 5.7$, $p < 0.0005$] – indicating significantly different changes in naming speed for the treated vs. untreated sets. The follow-up 2×3 ANOVAs confirmed that the effect of each therapy was significantly different from its control [Time Point × Set interactions were significant: RISP $F(2,36)=8.6$, $p=0.001$; SP $F(2,36)=3.9$, $p=0.03$]. A 2×3 ANOVA comparing the two treated sets indicated that there was a significant interaction between 'Time Point' and 'Treatment' [$F(2,36) = 3.2$, $p = .05$]. Whilst both treatments significantly reduced picture naming latencies between the baseline and both post-treatment assessments (1 week and 1 month), the pairwise t-tests showed that there was a trend for the RISP treatment to reduce RTs more than SP from baseline to the immediate assessment at Week 1 ($p = .101$) and, most strikingly, RISP was significantly more effective in maintaining the treatment effect in terms of quicker naming responses at the one month follow-up assessment ($p = .001$). In comparing the two untreated conditions, only the main effect of the 'Time Point' factor was significant ($F(2,36) = 3.23$, $p = .05$) – reflecting a small reduction in naming latencies across repeated assessments (presumably reflecting repetition priming). The main effect of 'Set' was not significant ($F(1,18) < 1$), nor was the interaction between 'Time Point' and 'Set' ($F(2,36) = 1.3$, $p = 0.28$).

Generalisation to connected speech: Word retrieval in composite picture descriptions

Again, exactly the same set of analyses were conducted on the target word retrieval data in the composite picture descriptions. The global 3×4 ANOVA indicated that there was a significant effect of the 'Time Point' factor [$F(2,38) = 87.8$, $p < .0005$], a main effect of 'Treatment' factor [$F(3,57) = 43.7$, $p < .0005$] and a highly significant interaction between 'Time Point' and 'Treatment' [$F(6,114) = 19.9$, $p < .0005$; (Figure 3c)] – indicating very different production of the target vs. untreated vocabulary in the patients' narratives before and after therapy. Directly comparing the two treatments (SP vs RISP), a 2×3 ANOVA indicated that there was a highly significant interaction between 'Time Point' and 'Treatment' [$F(2,38) = 19.6$, $p < .0005$]. The t-tests showed that the RISP effect on connected speech production was significantly stronger than SP both at the 1 week and 1 month post-treatment assessments (both $p < .0005$). Comparing each treatment to its control set, separately, we found significant 'Time Point' × 'Set' interactions for the RISP and SP sets [$F(2,38)=19.6$, $p<0.0005$; $F(2,38)=5.2$, $p=0.01$, respectively]. Thus, although there is a general clinical belief that standard therapy does not induce generalisation to connected speech, our newly-developed assessment was able to demonstrate that this is incorrect – there is, in fact, a small but significant generalisation to connected speech for SP both at one week and one month (though the effect was significantly smaller than for the RISP therapy – see above). Finally, the two untreated conditions were compared. The main effect of 'Time Point' was significant [$F(2,36) = 3.2$, $p = .05$], indicating a small improvement in target vocabulary production simply through repeated assessment, but neither the main effect of 'Set' [$F(1,18) < 1$] nor the interaction between 'Time Point' and 'Set' were significant [$F(2,36) = 1.3$, $p = 0.28$].

Content analysis of the connected speech samples

As well as exploring the generalization of trained vocabulary to the connected speech samples, it is also important to investigate the connected speech samples more generally. It is possible, for example, that improved vocabulary promotes connected speech more generally or that the improvement on the trained items comes at the cost of reduced performance on the untrained vocabulary. We examined the connected speech samples in terms of the total number of nouns produced (tokens), the number of unique nouns produced (types), nouns per minute, the type/token ratio (number of unique words divided by the total words), average word frequency and average imageability for the treated and untreated pictures.

The overall secondary effects on the patients' connected speech samples were entirely positive. Specifically, for the treated pictures, the speech samples including all items showed that significantly more unique items were produced after therapy compared to baseline (mean at 1 week = 103.6, mean at baseline = 85.4; t(18) = -2.30, p = 0.03). There was also a significant decrease in the average word frequency of the nouns used (mean at 1 week = 1.40, mean at baseline = 1.55; $t(18) = 4.21$, $p < 0.001$). There were no significant changes found in nouns per minute, type/token ratio, and average imageability rating. Importantly, there were no significant effects found in analyses of the untreated fifth picture, indicating that the improved connected speech samples did not reflect a non-specific effect of repeated assessment.

This first analysis included all items, including the target therapy items. Accordingly, we repeated the analysis to remove these items from consideration. In this second analysis, the increase in unique items from baseline to post therapy was no longer significant (mean at 1 week = 84.8, mean at baseline = 77.8; t(18) = -0.95, p = 0.3). The reduction in mean word frequency, however, was still significant (mean at 1 week = 1.48, mean at baseline = 1.58; t(18) = 2.86, p < 0.01).

Correlations with individual's background neuropsychological profile

Table 3 about here

Although there were significant and reliable therapy effects at the group level, the effect varied across individual patients. We performed correlations between the background neuropsychological profile (with respect to four principal neuropsychological components (see Table 3 for component loadings): phonological, semantic, executive, and speech quanta (fluency)) and the magnitude of the therapy effect (1 week vs. baseline performance, and 1 month (maintenance) vs. baseline performance) in order to reveal which aspects of the patients' profile were related to the therapy outcome. The PCA identified four components including phonological skill (50.9% variance), semantic ability (11.28% variance), executive ability (8.18% variance) and speech quanta (6.42% variance). In general, the phonological component loaded with repetition, naming and digit span tests, whereas the semantic component loaded with picture matching, camel and cactus and synonym judgement tests. The executive component loaded with Ravens coloured progressive matrices, Brixton spatial rule anticipation test and minimal pairs all of which are demanding tests. Finally, the measures of the amount of speech output component loaded on the fourth factor speech quanta. Overall, no correlations

were found between any of the components and the outcome on the standard therapy. For the RISP therapy, however, a significant negative correlation was found between the patients' phonological component score and the magnitude of their therapy effect, at both 1 week (r = -.55, p < .01) and 1 month (r = -.61, p < .005). This demonstrates that patients with the poorest phonological abilities showed the largest RISP benefit. As can be seen across the case-series (Figure 4), this negative correlation seems to reflect the fact that the RISP therapy was particularly beneficial leading to a ceiling effect for many of the milder patients (note that if patient JS with poor phonological abilities but a large therapy effect is removed, then the correlation is still significant).

Figure 4 about here

It was also possible to determine how each component correlated with the maintenance of the therapy effect (i.e., 1 month vs. 1 week performance). In this analysis, the maintenance of the RISP effect was found to correlate positively with performance on the executive tasks (r = .53, p < .01). Thus, the patients with better executive abilities exhibited the best therapy maintenance. No other correlations were significant.

Neural correlates of RISP

In order to determine the neural correlates of the RISP effect, we correlated each patient's therapy effect (1 week vs. baseline performance, and 1 month vs. baseline performance) with their T1-weighted MRI using voxel-based correlational methodology (VBCM: Tyler et al., 2005). This analysis revealed that patients with the greatest damage to the posterior superior temporal gyrus extending into the white matter of the inferior longitudinal fasciculus, showed the greatest RISP benefit both at 1 week and 1 month (height threshold p < .001, cluster corrected using FWE p < .05). This region is known to play an important role in phonological performance, as illustrated in Figure 5 whereby the RISP effect overlaps closely with the area related to the lesion correlate for the patients' phonological skill factor found previously by Halai et al. (2017) and thus aligns with the behavioural correlation between phonological ability and therapy effect noted above. It appears, therefore, that the RISP effect may relate particularly to the patients' phonological abilities. Finally, no voxels were found to correlate significantly with the RISP maintenance effect (1 month vs. 1 week performance).

Figure 5 about here

Discussion

Anomia is an immensely frustrating and disabling feature of aphasia, which is a common disorder post stroke (around 1/3 cases) and in other neurological conditions. Accordingly, it is important to establish effective interventions for remediating word-finding skills and generalising these improvements to patients' connected speech. Given the observation that fluent speech requires both quick and accurate word retrieval, we investigated and confirmed the novel hypothesis that a behavioural treatment, focussing on both speed and accuracy rather than accuracy alone (as is the case in standard methods), would generate greater improvements in both confrontation naming and also generalisation of this improved vocabulary to connected speech. A second key, novel feature of this study was that the interventions were not examined in isolation but we also investigated the neuropsychological and lesion correlates of treatment responsiveness. Although such analyses are a rarity in the literature to date (Abel et al., 2015), increasing our understanding about both the neuropsychological and lesion correlates of variable therapy success will be a critical step towards future neuroscience-led stratification of patients and choice of clinical pathways.

To address these questions, we developed a novel naming treatment that focussed on both speed and accuracy (RISP), which we compared to a standard accuracy-only treatment (SP). As expected, both treatments increased picture naming accuracy (assessed one week following the end of the intervention), which was largely retained at the one-month follow-up assessment even without maintenance practice. RISP was, however, significantly more effective than SP in promoting increased accuracy particularly at the important long-term follow-up assessment. The same pattern was found in naming speed – as intended, RISP was much more effective in speeding successful name retrieval and maintaining these improvements at follow-up assessment. Perhaps most importantly, we found that RISP generalised from naming individual target items into the patients' connected speech – a "holy grail" for speech and language therapy.

With regard to neuropsychological and neural correlates of therapy effects, we found a significant negative correlation for the RISP therapy between the patients' degree of phonological impairment and the magnitude of their therapy effect, both immediately after therapy and at follow up assessment. This initially somewhat counter-intuitive finding probably reflects that RISP appears to be an especially beneficial treatment, such that milder patients

show a resultant ceiling effect in their speech production assessment whereas the more severe patients can exhibit a much more dramatic improvement on the target items. This finding may also be consistent with the observation from Best and colleagues' (2013) meta-analysis that better treatment responsiveness was evident in participants classified as having relatively less semantic difficulties and greater phonological output deficits (note, our use of principal component analysis to extract the pattern of underlying language-cognitive deficits means that, over and above phonology per se, the potential additional influence of semantic, skills, speech fluency and cognitive-executive factors were already partialled out: see Butler et al., 2014; Halai et al., 2017). This behavioural correlate for the RISP therapy was also mirrored directly in the lesion correlate analysis: the RISP benefit was most evident in participants with the greatest damage to the posterior superior temporal gyrus extending into the white matter of the inferior longitudinal fasciculus. This region has been implicated in auditory-phonological processing not only through neuropsychological studies (Baldo et al., 2012; Robson et al., 2013, 2012) but also in fMRI explorations of healthy function (Hickok and Poeppel, 2004; Rauschecker and Scott, 2009; Warren and Griffiths, 2003). Finally, with regard to the long-term maintenance of the RISP treatment, follow-up performance correlated positively with cognitive-executive skills. Specifically, strong performances on neuropsychological assessments like the Brixton Rule Anticipation Test (Vordenberg et al., 2014) predict good longer-term responsiveness to anomia treatment in general, and RISP in particular. This may reflect the enhanced demands that RISP placed on participants in terms of cognitive flexibility, planning, problem-solving and speed of processing – consistent with the suggestion that both patients' degree of language impairment and remaining executive skill may be critical in recovery of function and therapy outcome (Brownsett et al., 2014; Geranmayeh et al., 2014; Lambon Ralph et al., 2010; Sharp et al., 2010).

Two different possible hypotheses can be made about the mechanisms underlying the speeded treatment effect, which can be tested in future investigations. The first, language-specific hypothesis is related to the aim of the RISP treatment to target both accuracy and speed. For optimally easy and efficient word retrieval, the language system requires precise representations that allow the target meaning to be converted to phonological and motor-speech representations (Lupker et al., 1997). Computational models of speech production and reading have repeatedly shown that as these representations and mappings are refined through learning, performance of models becomes both more accurate and more efficient (Ellis and Lambon

Ralph, 2000; Plaut et al., 1996). Accordingly, because the RISP treatment deliberately aims beyond accuracy to improve speed as well, the language representations and mappings may have been pressured not only to reform but also to be 'sharpened up' to become more precise. This also supports previous findings which indicated that naming speed is a significant yet often overlooked factor, not only in assessment but also in treatment tasks (McCall et al., 1997). Indeed, this hypothesis might also explain why, aside from speed, RISP led to significantly better naming accuracy than the accuracy-only focussed SP (following the fact that both speed and accuracy reflect the precision of the underlying language representations).

Another possible hypothesis accounting for the RISP effect is related to a domain-general, cognitive-executive mechanism (Geranmayeh et al., 2014; Lambon Ralph et al., 2010). Not only was the degree of treatment maintenance related to the patients' cognitive-executive skills, but all participants (irrespective of severity) reported RISP to be especially engaging and motivating. Thus, RISP may be much better than SP in engaging patients' executive and attentional skills, in addition to the speech production system, resulting in improved learning and retention. From a neurobiological perspective, increased motivation and reward-seeking behaviour has been being strongly associated with dopamine release (Fiorillo, 2013; Morita et al., 2013; Sharp et al., 2016) and dopamine has been associated with improved learning and therapy effects (Berthier and Pulvermüller, 2011; Gill and Leff, 2012). This observation speaks to the wider potential of 'gamification', that is utilising the dynamic and engaging aspects of commercial gaming software to ramp up the engagement required for rehabilitation tasks (Ferreira et al., 2014). Although based on a limited number of items in each condition, the current results suggest that there might be clinically-notable differences between the two therapy approaches, particularly at longer-term follow up. These indications from the current experimental exploration will need to be confirmed in larger-scale studies, including formal clinical trials.

References:

Abel, S., Weiller, C., Huber, W., Willmes, K., Specht, K., 2015. Therapy-induced brain reorganization patterns in aphasia. Brain 138, 1097–1112. https://doi.org/10.1093/brain/awv022

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Baldo, J.V., Katseff, S., Dronkers, N.F., 2012. Brain Regions Underlying Repetition and Auditory-Verbal Short-term Memory Deficits in Aphasia: Evidence from Voxel-based Lesion Symptom Mapping. Aphasiology 26, 338–354. https://doi.org/10.1080/02687038.2011.602391

Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F., 2003. Voxel-based lesion-symptom mapping. Nat. Neurosci. 6, 448–450. https://doi.org/10.1038/nn1050

Berthier, M.L., Pulvermüller, F., 2011. Neuroscience insights improve neurorehabilitation of poststroke aphasia. Nat. Rev. Neurol. 7, 86–97. https://doi.org/10.1038/nrneurol.2010.201

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., Howard, D., 2013. Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. Cortex 49, 2345–2357. https://doi.org/10.1016/j.cortex.2013.01.005

Bird, H., Lambon Ralph, M.A., Patterson, K., Hodges, J.R., 2000. The rise and fall of frequency and imageability: noun and verb production in semantic dementia. Brain Lang. 73, 17–49. https://doi.org/10.1006/brln.2000.2293

BNC Consortium, 2007. The British National Corpus, version 3 (BNC XML Edition). Oxford University Computing Services, Oxford, England.

Bozeat, S., Lambon Ralph, M.A., Patterson, K., Garrard, P., Hodges, J.R., 2000. Non-verbal semantic impairment in semantic dementia. Neuropsychologia 38, 1207–1215. https://doi.org/10.1016/S0028-3932(00)00034-8

Brownsett, S.L.E., Warren, J.E., Geranmayeh, F., Woodhead, Z., Leech, R., Wise, R.J.S., 2014. Cognitive control and its impact on recovery from aphasic stroke. Brain 137, 242–254. https://doi.org/10.1093/brain/awt289

Burgess, P., Shallice, T., 1997. Hayling and Brixton Tests. Pearson, Oxford, England.

Butler, R.A., Lambon Ralph, M.A., Woollams, A.M., 2014. Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures. Brain J. Neurol. 137, 3248–3266. https://doi.org/10.1093/brain/awu286

Conroy, P., Sage, K., Lambon Ralph, M.A.L., 2009. The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. Aphasiology 23, 707–730. https://doi.org/10.1080/02687030802165574

Crerar, M.A., 2004. Aphasia rehabilitation and the strange neglect of speed. Neuropsychol. Rehabil. 14, 173–206. https://doi.org/10.1080/09602010343000174

Ellis, A., Lambon Ralph, M.A., 2000. Age of Acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. JEP:LMC 26, 1103–1123.

Ferreira, C., Guimarães, V., Santos, A., Sousa, I., 2014. Gamification of Stroke Rehabilitation Exercises Using a Smartphone, in: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth '14. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, pp. 282–285. https://doi.org/10.4108/icst.pervasivehealth.2014.255326

Fiorillo, C.D., 2013. Two dimensions of value: dopamine neurons represent reward but not aversiveness. Science 341, 546–549. https://doi.org/10.1126/science.1238699

Geranmayeh, F., Brownsett, S.L.E., Wise, R.J.S., 2014. Task-induced brain activity in aphasic stroke patients: what is driving recovery? Brain 137, 2632–2648. https://doi.org/10.1093/brain/awu163

Gill, S.K., Leff, A.P., 2012. Dopaminergic therapy in aphasia. Aphasiology 28, 155–170. https://doi.org/10.1080/02687038.2013.802286

Goodglass, H., Kaplan, E., Barresi, B., 2000. Boston Diagnostic Aphasia Examination - Third Edition (BDAE-3). Pearsons.

Halai, A.D., Woollams, A.M., Lambon Ralph, M.A., 2017. Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. Cortex, Is a "single" brain model sufficient? 86, 275–289. https://doi.org/10.1016/j.cortex.2016.04.016

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Hodgson, C., Lambon Ralph, M., 2008. Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. Brain Lang. 104, 89–101.

Jefferies, E., Patterson, K., Jones, R.W., Lambon Ralph, M.A., 2009. Comprehension of concrete and abstract words in semantic dementia. Neuropsychology 23, 492–499. https://doi.org/10.1037/a0015452

Kay, J., Lesser, R., Coltheart, M., 1996. Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. Aphasiology 10, 159–180. https://doi.org/10.1080/02687039608248403

Laine, M., Martin, N., 2006. Anomia: Theoretical and clinical aspects. Psychology Press, Hove, England.

Lambon Ralph, M.A., Snell, C., Fillingham, J.K., Conroy, P., Sage, K., 2010. Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. Neuropsychol. Rehabil. 20, 289–305. https://doi.org/10.1080/09602010903237875

Levelt, W., 1989. Speaking: from intention to articulation. MIT Press, Cambridge, Mass.

Lupker, S.J., Brown, P., Colombo, L., 1997. Strategic control in a naming task: Changing routes or changing deadlines? J. Exp. Psychol. Learn. Mem. Cogn. 23, 570–590.

Maendl, L., 1998. Word finding difficulties in aphasia. University of York.

McCall, D., Cox, D.M., Shelton, J.R., Weinrich, M., 1997. The influence of syntactic and semantic information on picture-naming performance in aphasic patients. Aphasiology 11, 581–600. https://doi.org/10.1080/02687039708248491

Morita, K., Morishima, M., Sakai, K., Kawaguchi, Y., 2013. Dopaminergic control of motivation and reinforcement learning: a closed-circuit account for reward-oriented behavior. J. Neurosci. Off. J. Soc. Neurosci. 33, 8866–8890. https://doi.org/10.1523/JNEUROSCI.4614-12.2013

Nickels, L., 2002. Therapy for naming disorders: Revisiting, revising, and reviewing. Aphasiology 16, 935–979. https://doi.org/10.1080/02687030244000563

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K., 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. Psychol. Rev. 103, 56–115.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. https://doi.org/10.1038/nn.2331

Raven, J., 1962. Raven's Coloured Progressive Matrices (CPM). Pearson, Oxford, England.

Robson, H., Grube, M., Lambon Ralph, M.A., Griffiths, T.D., Sage, K., 2013. Fundamental deficits of auditory perception in Wernicke's aphasia. Cortex J. Devoted Study Nerv. Syst. Behav. 49, 1808–1822. https://doi.org/10.1016/j.cortex.2012.11.012

Robson, H., Sage, K., Ralph, M.A.L., 2012. Wernicke's aphasia reflects a combination of acoustic-phonological and semantic control deficits: a case-series comparison of Wernicke's aphasia, semantic dementia and semantic aphasia. Neuropsychologia 50, 266–275. https://doi.org/10.1016/j.neuropsychologia.2011.11.021

Schneider, W., Eschman, A., Zuccolotto, A., 2002. E-Prime Computer Software and Manual. Psychology Software Tools Inc., Pittsburg, PA.
Seghier, M.L., Ramlackhansingh, A., Crinion, J., Leff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. Neuroimage 41, 1253–1266. https://doi.org/10.1016/j.neuroimage.2008.03.028
Sharp, D.J., Turkheimer, F.E., Bose, S.K., Scott, S.K., Wise, R.J.S., 2010. Increased frontoparietal integration after stroke and cognitive recovery. Ann. Neurol. 68, 753–756. https://doi.org/10.1002/ana.21866
Sharp, M.E., Foerde, K., Daw, N.D., Shohamy, D., 2016. Dopamine selectively remediates "model-based" reward learning: a computational approach. Brain 139, 355–364. https://doi.org/10.1093/brain/awv347
Swinburn, K., Porter, G., Howard, D., 2004. Comprehensive Aphasia Test. Psychology Press, Hove, England.
The Stroke Association (UK), 2016. State of the Nation Stroke statistics January 2016.
Tyler, L.K., Marslen-Wilson, W., Stamatakis, E.A., 2005. Dissociating neuro-cognitive component processes: voxel-based correlational methodology. Neuropsychologia 43, 771–778. https://doi.org/10.1016/j.neuropsychologia.2004.07.020
Van Casteren, M., Davis, M.H., 2007. Match: a program to assist in matching the conditions of factorial experiments. Behav. Res. Methods 39, 973–978.
Vitkovitch, M., Humphreys, G.W., 1991. Perseverant responding in speeded naming of pictures: It's in the links. J. Exp. Psychol. Learn. Mem. Cogn. 17, 664–680. https://doi.org/10.1037/0278-7393.17.4.664
Vordenberg, J.A., Barrett, J.J., Doninger, N.A., Contardo, C.P., Ozoude, K.A., 2014. Application of the Brixton Spatial Anticipation Test in Stroke: Ecological Validity and Performance Characteristics. Clin. Neuropsychol. 28, 300–316. https://doi.org/10.1080/13854046.2014.881555
Warren, J.D., Griffiths, T.D., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J. Neurosci. Off. J. Soc. Neurosci. 23, 5799–5804.
Wechsler, D., 1945. A Standardized Memory Scale for Clinical Use. J. Psychol. 19, 87–95. https://doi.org/10.1080/00223980.1945.9917223
Wisenburn, B., Mahoney, K., 2009. A meta-analysis of word-finding treatments for aphasia. Aphasiology 23, 1338–1352. https://doi.org/10.1080/02687030902732745

List of Figures and Tables

Figures

Supplementary Figure

Tables

Supplementary Tables